

# 基于VuFind的OA及自建资源发现系统

暨南大学图书馆

张春晓

- ▶ 项目背景
- ▶ 软件平台的选择
- ▶ 项目实施进度
- ▶ 所遇到的问题
- ▶ 创新点
- ▶ 未来计划



# 项目背景

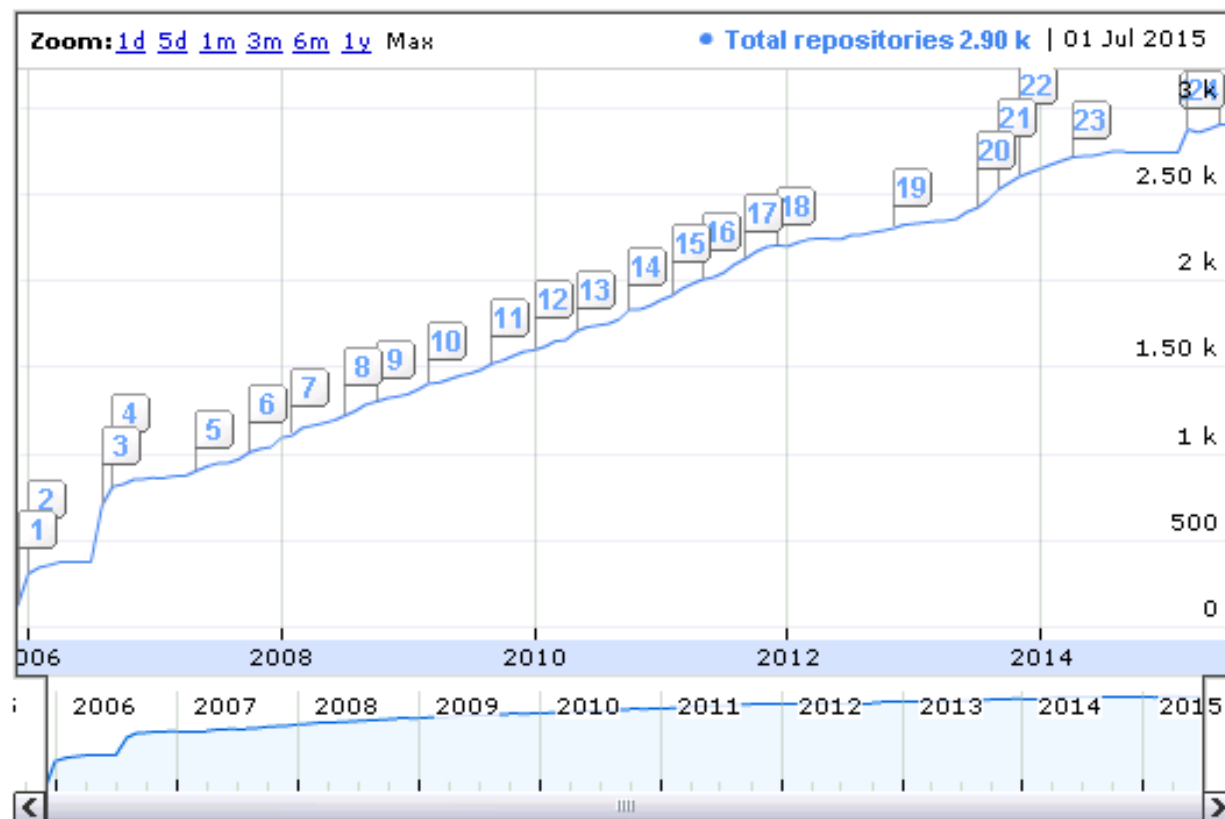


# 项目背景（1）：开放获取资源增长迅速

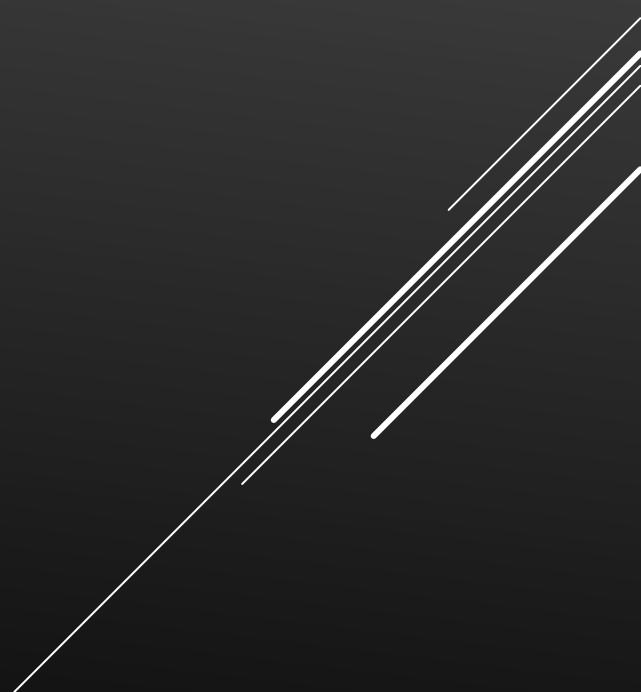
- ▶ 随着互联网的发展，开放获取已经成为一种全新的学术传播机制。
- ▶ 截至2015年7月5日，DOAJ (Directory of Open Access Journal, 开放存取期刊目录) 已经收录来自134个国家的10,363种期刊，1,888,124篇文章。
- ▶ 到目前为止，在OpenDOAR（由英国的诺丁汉大学和瑞典的伦德大学图书馆于2005年2月共同创建的开放获取机构资源库、学科资源库目录检索系统）注册的OA仓库已达2902个。
- ▶ 2010年，由中国科学院主管的中国科技期刊开放获取平台(China Open Access Journals, COAJ)正式上线运行。截至2015年7月5日，已经收录644种期刊，1417358篇论文。
- ▶ 各高校纷纷建设自己的机构知识库。

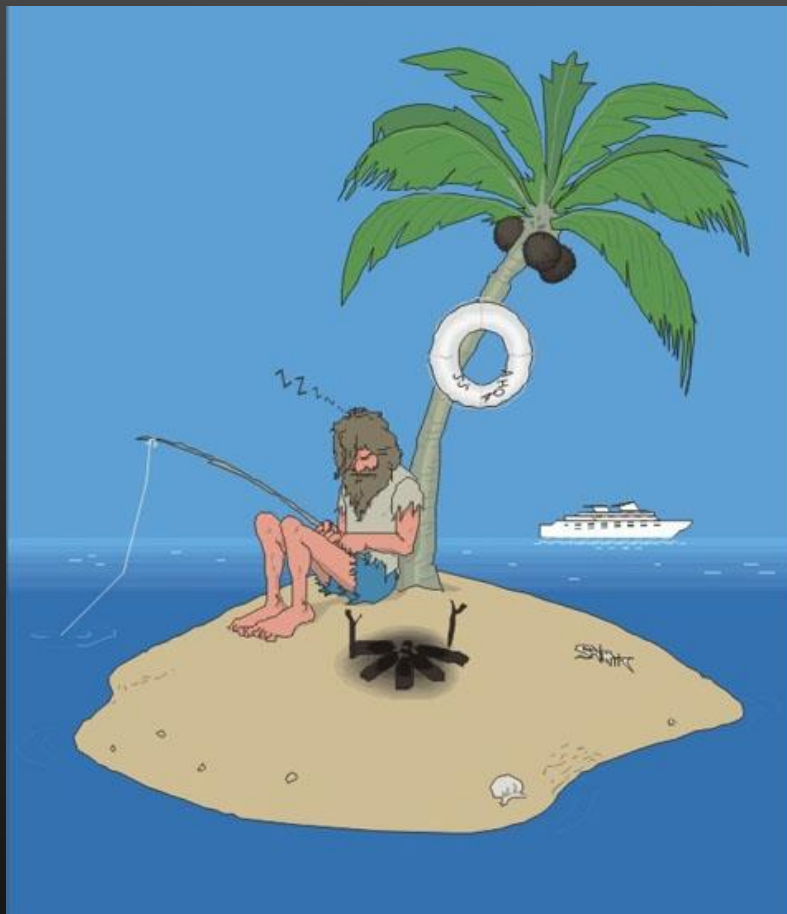
## Growth of the OpenDOAR Database - Worldwide

OpenDOAR - 05-Jul-2015



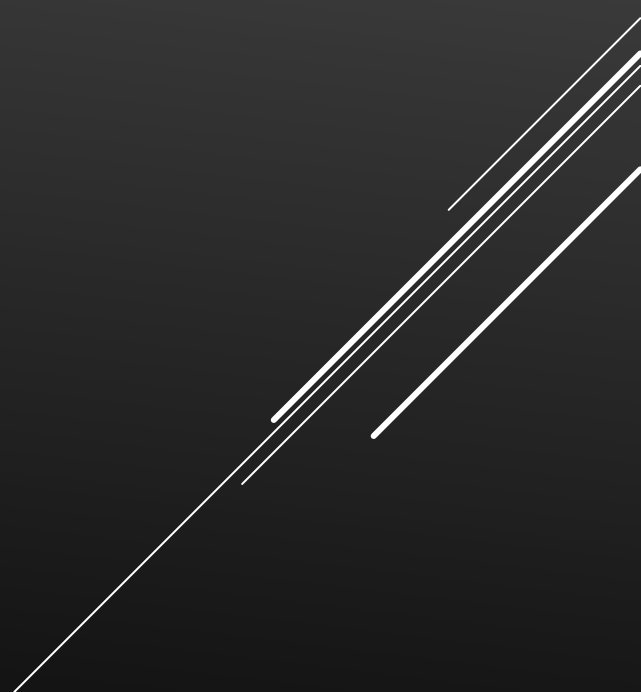
## 项目背景（2）：自建资源日益增多

- ▶ 博硕士学位论文数据库
  - ▶ 馆藏图书全文数据库
  - ▶ 华侨华人书刊全文数据库
  - ▶ 华侨华人人物数据库
  - ▶ 华侨华人图片数据库
  - ▶ 侨务信息数据库
  - ▶ 开放课程视频数据库
  - ▶ .....
- 



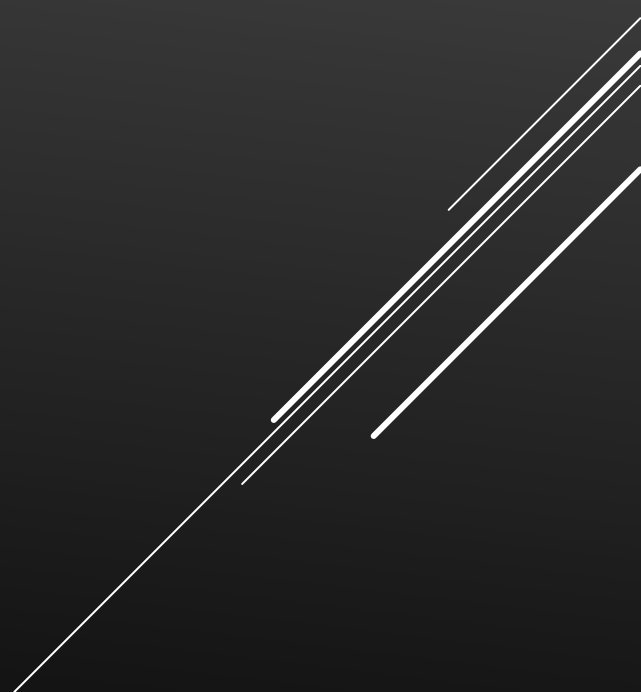
信息孤岛现象严重

# 软件平台的选择





# 选择依据

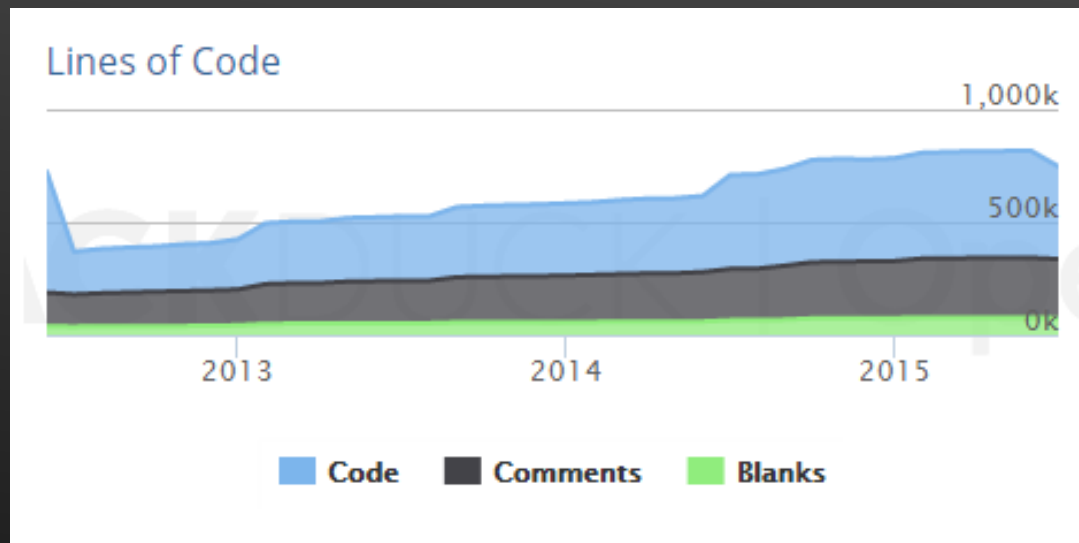
- ▶ 必须是开放的，能够自主选择收录哪些数据。
  - ▶ 必须实现OAI-PMH协议，即能够通过OAI协议收割元数据。
  - ▶ 必须能够存储各种规格的XML元数据。
  - ▶ 必须能够使得用户提交各种非XML元数据。
  - ▶ 必须能够提供高效的搜索性能和分面搜索功能。
  - ▶ 必须能够与第三方系统，例如用户认证系统集成。
- 
- A series of several parallel white lines of varying lengths and slopes, located in the bottom right corner of the slide, serving as a decorative element.

# 选择之一：LibraryFind

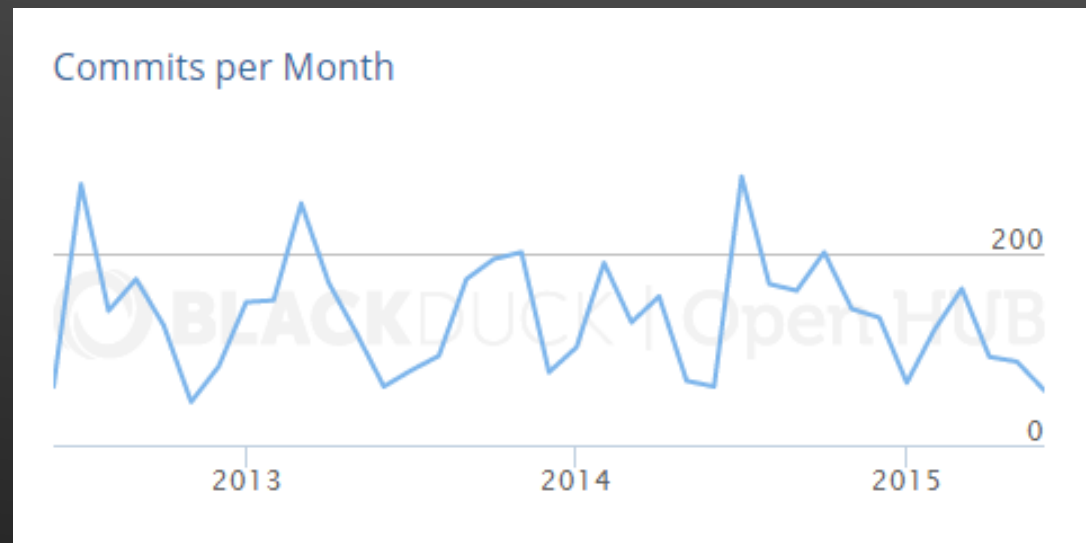
- ▶ 由美国俄勒冈州立大学图书馆开发。
- ▶ 设计目标为易于使用、安装和维护的图书馆资源整合查询系统，用于可以通过2次点击获取任何资源：一次点击用来查询，一次点击用来获取。
- ▶ 采用Ruby On Rails开发。
- ▶ 采用MySQL作为后台数据库。
- ▶ 提供元数据收割、索引以及前台页面。
- ▶ 目前已经停止维护。官方网站已经无法访问，只能在GitHub上面下载到3年前的代码。

## 选择之二：VuFind

- ▶ 由美国维拉诺瓦大学法尔维纪念图书馆于2005年10月开始开发，并获得了2008年第三届梅隆技术协作奖。
- ▶ 大部分采用PHP开发，应用了Zend Framework框架，实现了MVC三层架构。元数据库导入采用Linux Shell脚本结合XML样式表实现。
- ▶ 采用MySQL作为后台数据库。
- ▶ 采用Apache Solr作为后端索引和搜索引擎。
- ▶ 提供OAI元数据收割、索引、全文搜索、关键字高亮显示以及分面搜索功能。
- ▶ 此外用户还可以通过Solr API来提交其他类型的元数据。
- ▶ 提供移动设备界面。
- ▶ 开发活跃，2015年5月25日刚刚发布了版本2.4.1。



## 代码增长量



## 代码提交次数变化趋势

展示层

用户界面

应用层

VuFind

Application Controller

搜索

浏览

记录

OAI

其他

数据层

Solr

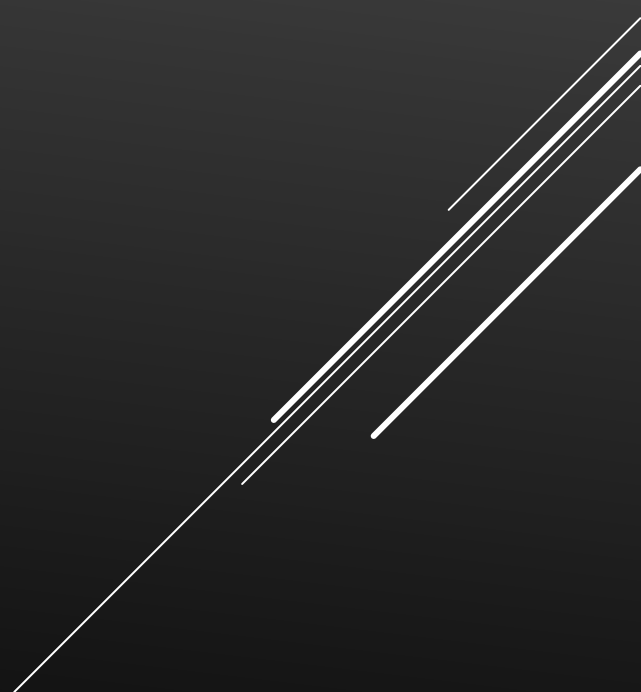
MySQL

ILS

other

VuFind三层架构

# 项目实施进度



# 项目实施进度

- ▶ 2014年10月开始筹备，测试和选择软件平台。
- ▶ 2015年3月份完成VuFind的安装调试。
- ▶ 2015年5月份完成网页模板的初步修改。
- ▶ 2015年6月份导入馆藏图书元数据150多万条；导入DOAJ论文元数据100多万条；导入中国科学院机构库20多万条；厦门大学学术典藏库81049条。
- ▶ 2015年6月实现用户认证与图书馆用户中心的集成。
- ▶ 2015年6月，命名为“四海学术搜索”，来自于《尚书·禹贡》“东渐于海，西被于流沙，朔南暨声教，讫于四海”。



四海学术搜索首页





关键字自动完成

检索结果 - cell

202.116.13.203/vufind/Search/Results?lookfor=cell&type=AllFields

四海学术搜索  
Search Discover & Share

0 items 登录 语言

cell 全文检索 检索 高级检索

检索: cell

搜索替代词:  
cell » well

显示 1 - 20 of 55,761 用于搜索: 'cell', 查询时间: 0.15s 排序 相关性排序

☐ 选择页 | 与选定的: 电子邮件 导出 打印 保存 添加到书包

1  Expressions of Fas/DcR3 and RGD-FasL mediated apoptosis in pituitary adenomas  
由 Zhuang, GH  
出版日期: 2013  
子库: Xiamen University Institutional Repository  
主题: "...CELLS..."  
获取全文  
Online

2  DNA Aptamer-Mediated Cell Targeting  
由 Xiong, Xiangling  
出版日期: 2013  
子库: Xiamen University Institutional Repository  
主题: "...cell modification..."  
获取全文  
Online

正在等待 202.116.13.203 的响应...

缩小搜索

格式

|         |        |
|---------|--------|
| Online  | 42,701 |
| Article | 12,754 |
| E-book  | 306    |

作者

|                    |    |
|--------------------|----|
| Li                 | 29 |
| 张华民                | 29 |
| Wang               | 27 |
| Zhang              | 26 |
| Liu                | 21 |
| Carlo Alberto Redi | 19 |
| 更多 ...             |    |

语言

|         |        |
|---------|--------|
| English | 32,301 |
|---------|--------|

分面检索功能

二维码



全文链接

网址

<http://dspace.xmu.edu.cn:8080/dspace/handle/2288/71341>



引用



发送短信



推荐此记录



导出记录



加到收藏夹



添加到书包

## Expressions of Fas/DcR3 and RGD-FasL mediated apoptosis in pituitary adenomas

Zhongshan Hospital of Xiamen University; Health Bureau of Xiamen, China

主要作者:

Zhuang,GH

其他作者:

Chen,LK, Tian,XH, Li,WZ, Agarwal,A, 庄国洪

格式:

Online

语言:

en\_US

出版日期: :

Neurology India, 2009,57(1):28-30 2013

主题:

CELLS

在线阅读:

<http://dspace.xmu.edu.cn:8080/dspace/handle/2288/71341>

标签:

没有标签, 成为第一个标记此 [添加标签](#) 记录!

### 相似书籍

RGD-FasL induces apoptosis of pituitary adenoma cells

由: Zhuang,GH

出版日期: : (2013)

Correlation between expression of DcR3 on tumor cells and sensitivity to FasL

由: Zhuang,GH

出版日期: : (2013)

RGD-FasL Induces Apoptosis in Hepatocellular Carcinoma

由: Liu, Zhongchen

出版日期: : (2013)

RGD-FasL Induces Apoptosis in Hepatocellular Carcinoma

由: Wang,J

出版日期: : (2013)

Induction of tumor cell apoptosis via Fas/DR5 (vol 3 pg 467, 2006)

由: Zhuang,GH

出版日期: : (2013)

相似文献

详细记录页面



移动设备



PC电脑

响应式页面设计，满足不同设备浏览的需求

所遇到的问题



# 所遇到的问题（1）

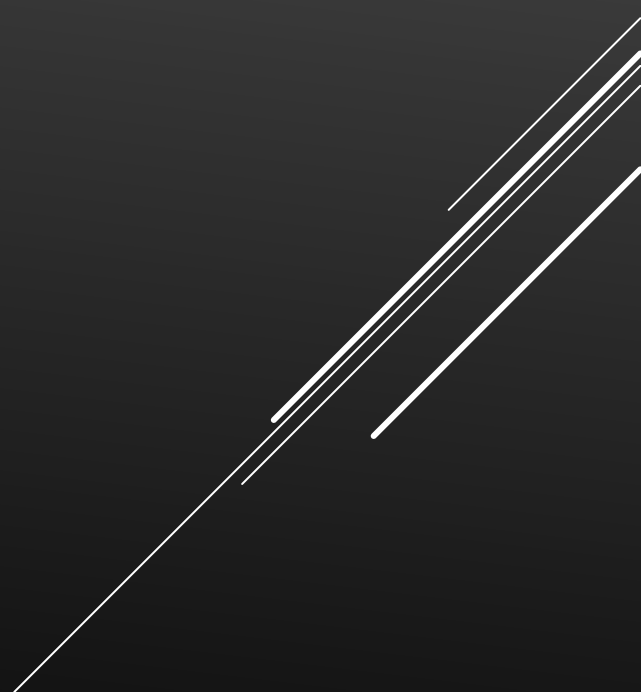
- ▶ 对于大的OA仓储，元数据收割时会经常由于连接超时而中断。

解决方法：采用Linux Cron计划任务，定时判断收割进程是否存在。如果进程不存在，则重新启动收割进程。

缺陷：需要从头开始重新收割，浪费大量时间。

## 所遇到的问题 (2)

- ▶ 尽管每个OA仓储都提供DC元数据格式，但是实际上有些字段并不统一，例如全文链接放在relation字段里面，DOAJ将全文链接放在dc:identifier字段里面，需要编写复杂的XML样式表来进行转换，对数据收割者要求较高。



```
<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <dc:title>Croatian Economic Survey</dc:title>
  <dc:identifier>1330-4860</dc:identifier>
  <dc:identifier>1846-3878</dc:identifier>
  <dc:identifier>
    http://doaj.org/search?source=%7B%22query%22%3A%7B%22bool%22%3A%7B%22must%22%3A%5B%7B%22term%22%3A%7B%22id%22%3A%229dfe4879964541069dff9bb6d15b65a2%22%7D%7D%5D%7D%7D%7D
  </dc:identifier>
  <dc:subject>economics</dc:subject>
  <dc:subject>post-socialist Europe</dc:subject>
  <dc:subject>comparative economics</dc:subject>
  <dc:subject>policy formulation</dc:subject>
  <dc:language>English</dc:language>
  <dc:relation>
    http://www.eizg.hr/en-US/Croatian-Economic-Survey-26.aspx
  </dc:relation>
  <dc:publisher>The Institute of Economics, Zagreb</dc:publisher>
  <dc:relation>
    http://www.eizg.hr/en-US/Croatian-Economic-Survey-26.aspx
  </dc:relation>
  <dc:publisher>The Institute of Economics, Zagreb</dc:publisher>
  <dc:date>2013-11-06T15:18:52Z</dc:date>
  <dc:type>journal</dc:type>
```

来自DOAJ的元数据



```

▼<metadata>
▼<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
▼<dc:title>
Epidemiology of Rifampin ADP-Ribosyltransferase (arr-2) and Metallo- $\beta$ -Lactamase (blaIMP-4) Gene Cassettes in Class 1 Integrons in Acinetobacter Strains Isolated from
2000
</dc:title>
<dc:creator>Houang, Elizabeth T. S.</dc:creator>
<dc:creator>Chu, Yiu-Wai</dc:creator>
<dc:creator>Lo, Wai-Sing</dc:creator>
<dc:creator>Chu, Ka-Yi</dc:creator>
<dc:creator>Cheng, Augustine F. B.</dc:creator>
<dc:subject>Mechanisms of Resistance</dc:subject>
<dc:description/>
<dc:publisher>American Society for Microbiology</dc:publisher>
<dc:date>2003-04</dc:date>
▼<dc:identifier>
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC152494/
</dc:identifier>
<dc:identifier>http://www.ncbi.nlm.nih.gov/pubmed/12654674</dc:identifier>
<dc:identifier>http://dx.doi.org/10.1128/AAC.47.4.1382-1390.2003</dc:identifier>
<dc:type>Text</dc:type>
<dc:language>en</dc:language>
▼<dc:rights>
Copyright © 2003, American Society for Microbiology
</dc:rights>
</oai_dc:dc>
</metadata>

```

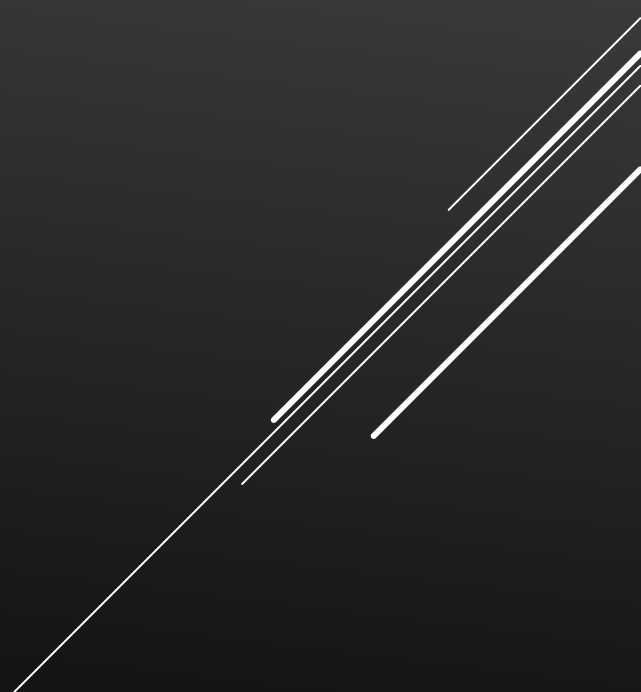
来自PUBMED的元数据

```
▶ <xsl:stylesheet version="1.0">
▶   <xsl:output method="xml" indent="yes" encoding="utf-8"/>
▶   <xsl:param name="collection">DOAJ</xsl:param>
▶   <xsl:template match="doaj:record">
▶     <add>
▶       <!-- AUTHOR -->
▶       <xsl:if test="//doaj:authors/doaj:author/doaj:name">
▶         <xsl:for-each select="//doaj:authors/doaj:author/doaj:name">
▶           <xsl:if test="normalize-space()">
▶             <xsl:if test="position()=1">
▶               <field name="author">
▶                 <xsl:value-of select="normalize-space()"/>
▶               </field>
▶               <field name="author-letter">
▶                 <xsl:value-of select="normalize-space()"/>
▶               </field>
▶             </xsl:if>
▶           </xsl:if>
▶         </xsl:for-each>
▶       </xsl:if>
▶     </add>
▶   </xsl:template>
▶ </xsl:stylesheet>
```

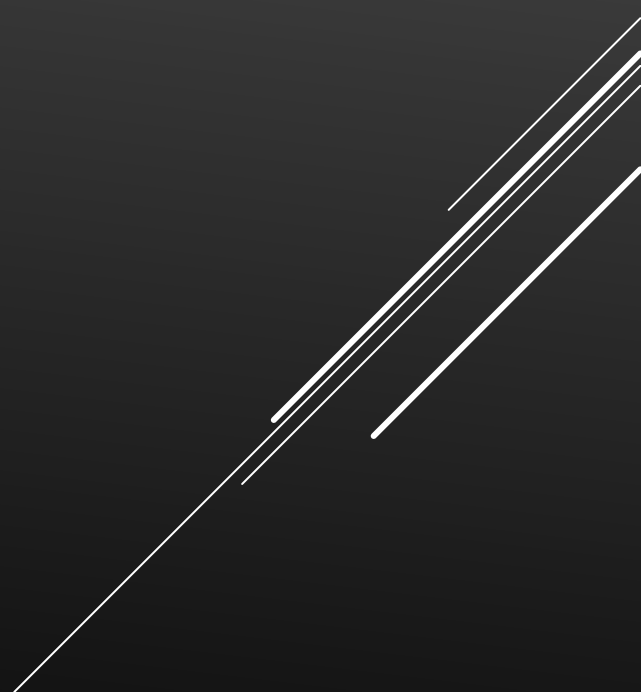
DOAJ元数据转换样式表（部分）

## 所遇到的问题 (2)


- ▶ 二次开发难度大，需要开发者熟练掌握PHP、Zend Framework、Apache Solr、Java以及Shell脚本等语言以及开发框架。



创新点



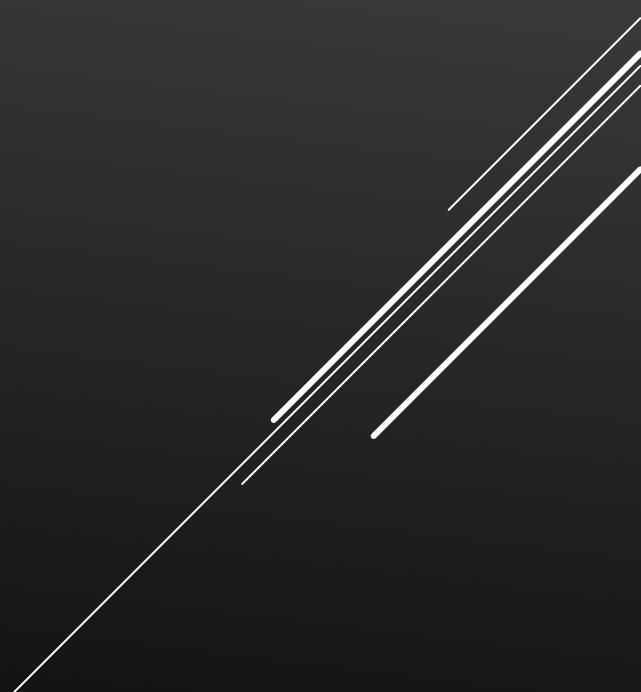
# 创新点

- ▶ 在国内首次引入VuFind，并且用之于实践。
  - ▶ 在一定程度上解决了OA资源以及自建资源的信息孤岛问题，提高了这些资源的利用率。
  - ▶ 采用响应式网页设计，以满足不同设备访问资源的需求。
- 
- A series of white lines of varying lengths and orientations are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

未来计划



# 未来计划

- ▶ 继续搜集数据源，充实元数据。
  - ▶ 改善元数据收割模块的功能，提供更加完善的收割功能。
  - ▶ 加强与第三方系统的整合，例如本馆开发的国际论文资讯网以及OPAC等。
  - ▶ 开发更加完善、全面的统计分析功能。
- 
- A series of white lines of varying lengths and orientations are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

欢迎批评指正，谢谢！

